Python, Transformers, pyTorch, Tensorflow, pandas, numpy, C++, Kaldi, scikit-learn, OpenCV, Docker, Kubernetes, AWS,

smg2280@columbia.edu & LinkedIn & GitHub & Google Scholar & Website

EDUCATION

Columbia University (MS in Computer Science)

Relevant Coursework: NLP, Conversational AI (Dialog Systems), Speech Recognition, Adv Spoken language processing, Deep learning for Computer Vision, AI, ML and Climate.

SAYALI GHODEKAR

Pune Institute of Computer Technology (BE in Computer Engineering)

Relevant Coursework: Machine Learning, Data Analytics, AI and Robotics, Soft Computing and Optimization, Engineering Mathematics, Cloud Computing, High Performance Computing. GPA: 3.97/4

WORK EXPERIENCE

RingCentral Inc, Machine learning engineer (Conversational AI Team)

- Created innovative Machine learning strategies for NLP problems, leading to integration of AI-based services in RingCentral's video communication platform for 300K businesses worldwide.
- Built, deployed and scaled abstractive summarization service to extract key insights from conversations, decreasing system latency. Implemented data collection, models and client libraries, dockerization, and Kubernetes deployments.
- Improved qualitative performance of post-meeting tools through LLM text-generation services including Coreference-Resolver, question-generator, text-paraphraser, and multilingual summarizer.

Consumer Reports, Data Science Intern

- Created a pre-processing pipeline for cleaning noisy data of 50K brands scraped from CR's website using NLP techniques.
- Collaborated with the Product Safety team to design early warning system that integrates CR's 5 internal data channels using Twitter API and Transformers (BERT).

DeepAffects (Conversation-Intelligence Startup), Machine learning engineer

- Developed and scaled DeepAffect's abstractive summarization API, leading to a 10% increase in model performance. Collaborated with quality team for continuous testing, processing over 1M+ minutes of audio into summaries.
- Enhanced DeepAffect's conversation metrics stack through dialogue act tagging, intent classification, question and answering systems, reducing time to market these APIs by 75%.
- Proposed and implemented a topic segmentation algorithm for multi-turn dialogue data, demonstrating an improvement in model performance across all NLP pipelines.

RESEARCH

Columbia NLP Lab

Advised by Prof. Zhou Yu (in collaboration with Prof. Shinji Watanabe, CMU LTI)

- Enhanced Edubot with spoken-language learning feature. Built ASR and TTS models for recognizing accented speech of non-native English speakers.
- Built ML model for automated evaluation of spoken-language proficiency using lexical and fluency-based metrics.

Center for Indian Language Technologies, IIT Bombay

PI: Prof. Pushpak Bhattacharyya

- Designed experiments for Cognate Detection with Siamese neural networks, evaluating word embedding models and beating baseline by 71%.
- Collaborated with Cognitive NLP group to derive gaze features for cognate identification. Reduced annotation cost by leveraging predictive models, demonstrating a 12% improvement over baseline.
- Published and presented cognate study findings in ACM CoDS-COMAD 2021 [link] and EACL 2021 [link]. Awarded best paper honorable mention at EACL 2021.

TEACHING

Natural language processing, Fall 2021 and Fall 2022

Javascript, PHP, MySQL, MongoDB, PowerShell.

Teaching Assistant for Graduate-level course at Columbia CS (Responsibilities: Weekly recitations, Office hours and Grading)

ACADEMIC PROJECTS

Automatic Speech Recognition System for Code-Switched speech

• Built an ASR system for recognizing Code-switched text for the low-resource Hindi-English language pair using the Kaldi toolkit. Created my self-recorded dataset and achived a WER of 36 on the test set.

TECHNICAL SKILLS

Aug 2022-

June 2019 - Dec 2019

Dec 2020 - July 2021

May 2022 - Aug 2022

Jan 2020 - Dec 2020

June 2019

Expected Dec 2022

New York. NY

Sep 2021 - Dec 2021